

IT Real-Time training that work for your career.

PROVIDED TRAINING FOR THOUSANDS OF STUDENTS.

SUBJECT, MATERIAL & VIDEOS

**VIDEOS
PROVIDED**

MS-BI

DETAILED THEORY, FAQs, SCENARIOS, & REAL-TIME

MATERIAL

MSBI

IS, AS, RS & MDS

POWER BI

SERVER, DESKTOP & DAX

Trainings:

CLASS ROOM



ONLINE

FAST TRACK



ONE ON ONE

PROJECT TRAINING

Address:

Flat No: 506/B
Nilgiri Block
Aditya Enclave
Mytrivanam Area
Hyderabad.

Website & Blog

www.vinaytechhouse.com

www.msbivinay.blogspot.in

Contact Information

+91 9573168449

040 66638869



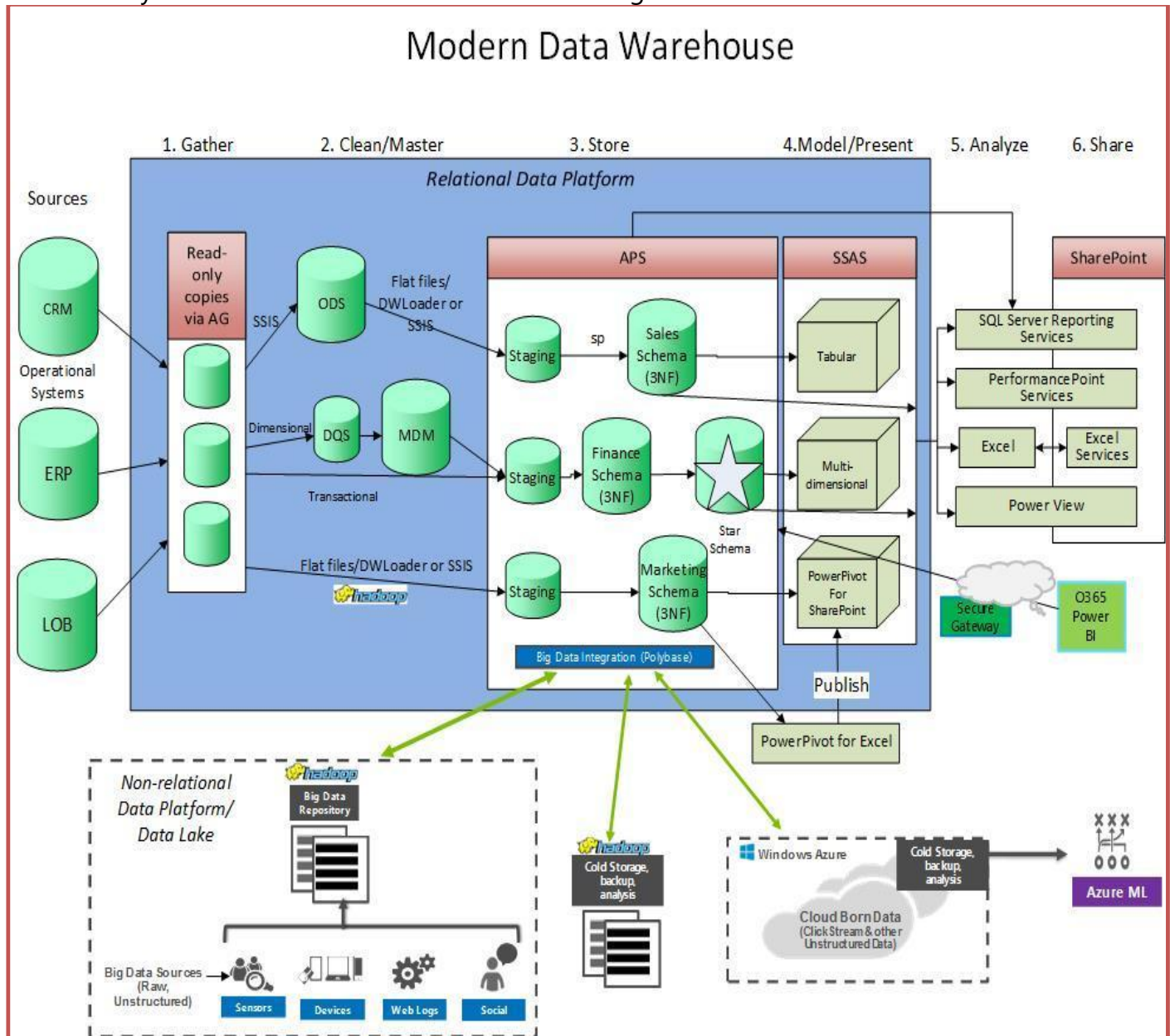
WE'VE WORKED WITH A DIVERSE CUSTOMER BASE. HOW CAN WE HELP YOU?

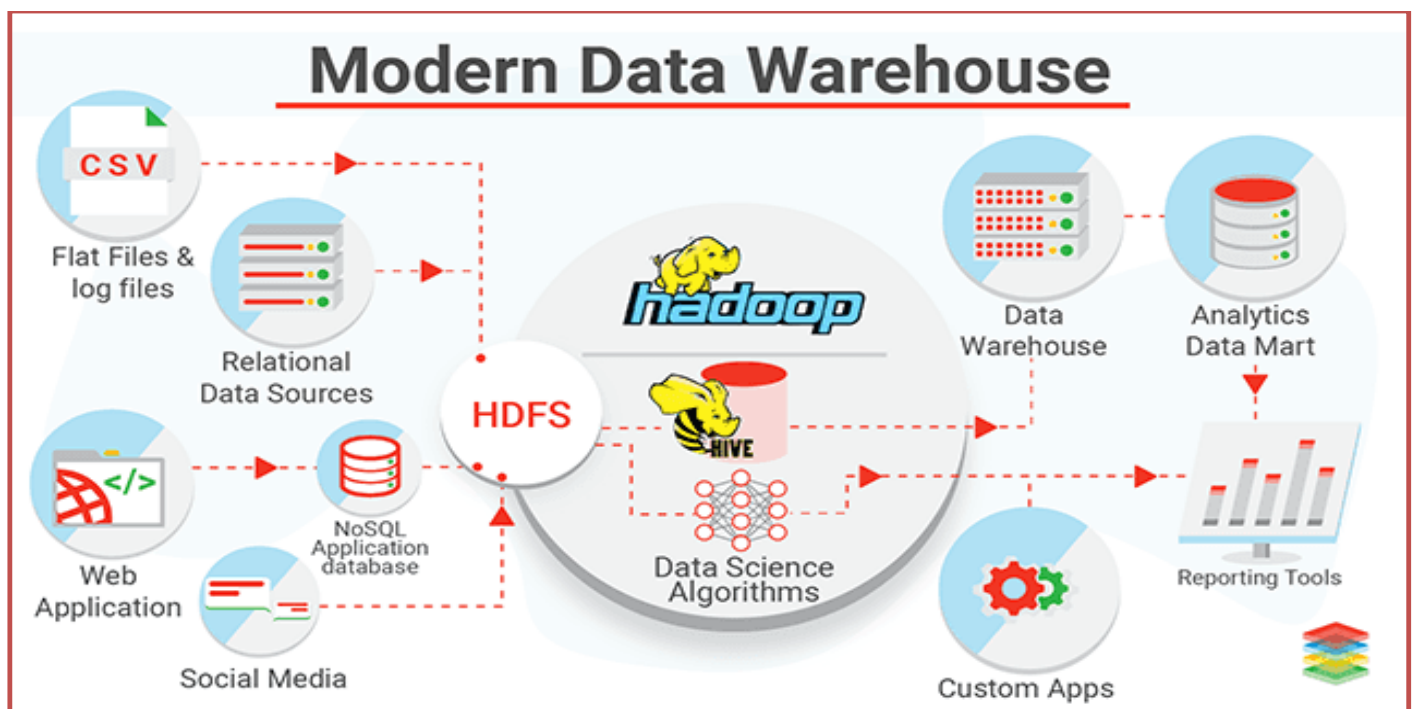
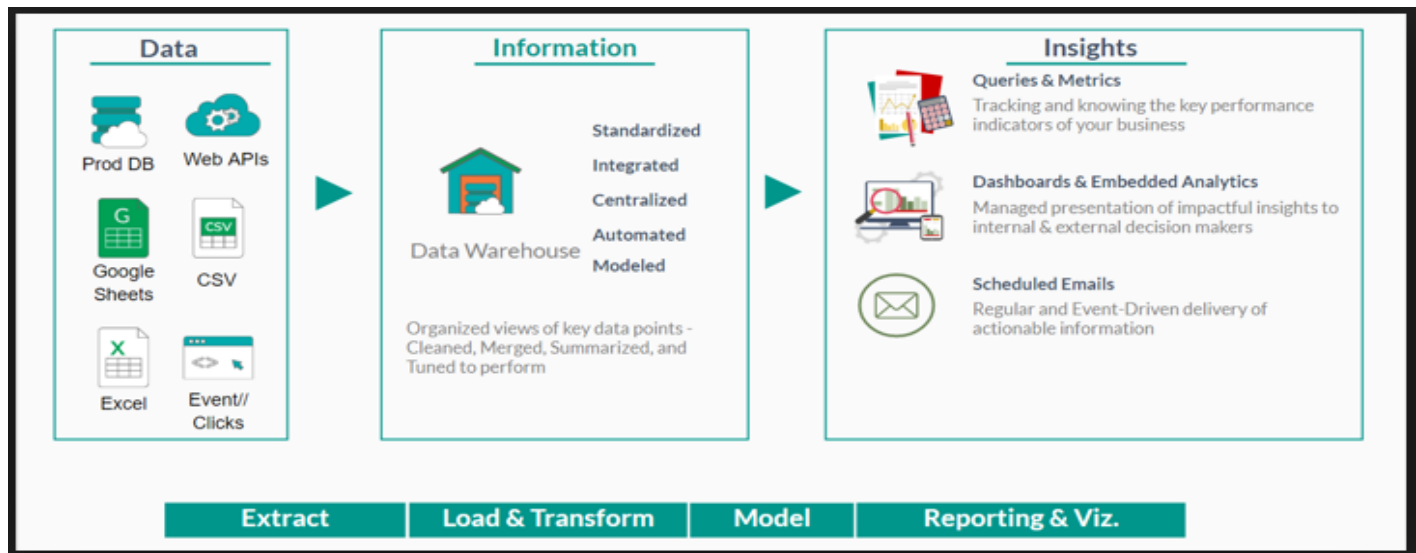
IT Training, Support and Consulting.

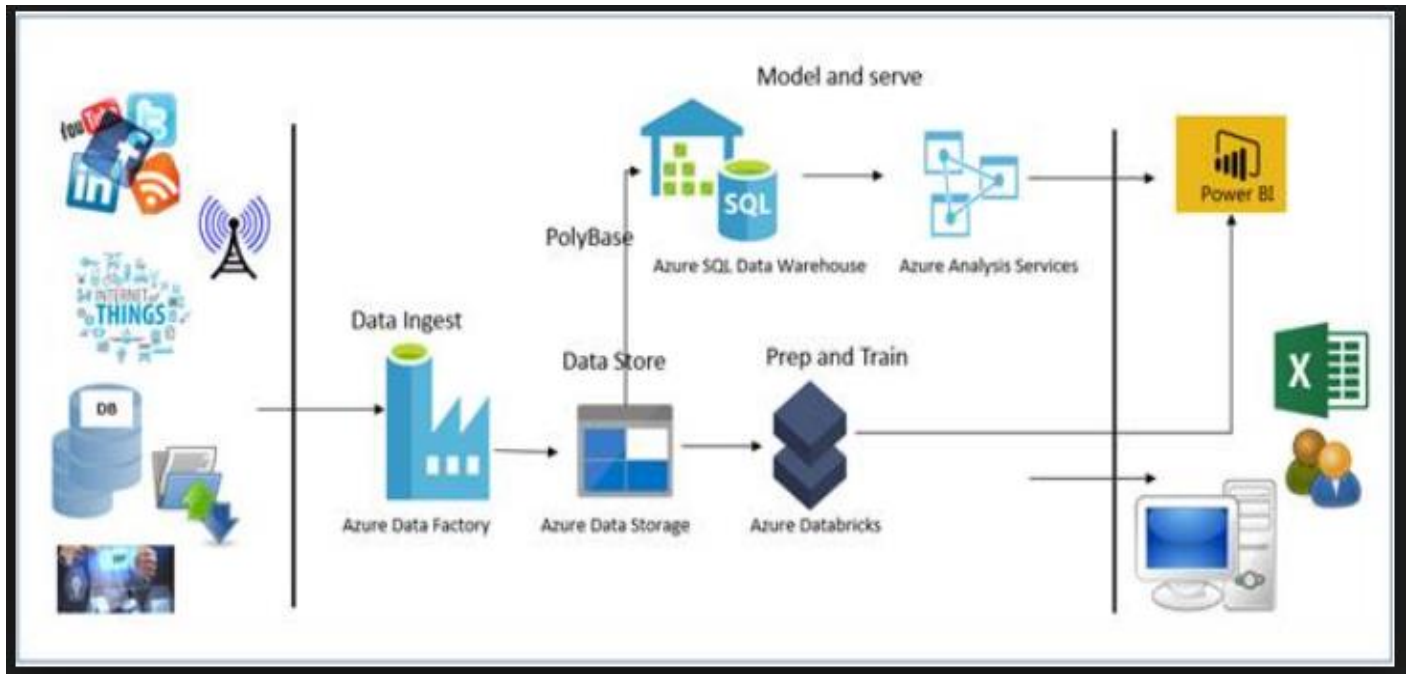
DATAWAREHOUSE [MODERN & LEGACY] FUNDAMENTALS

AUDIENCE

- Every Professional in DWH and Business Intelligence







DATA

DATA is composed of observable and recordable facts that are often found in operational and transactional system.

INFORMATION

Information is an integrated collection of facts and is used as the basis for decision making. In a data warehouse environment, **data** only comes to have value to end-users when it is organized and presented as **information**

OLTP (ONLINE TRANSACTION PROCESSING)

OLTP is an abbreviation of ONLINE-TRANSACTION PROCESSING. This system is an Applicable Application that modifies data the instance it receives and has a large number of concurrent users.

ODS (OPERATIONAL DATA STORE)

It stores near real time data from transactional processing applications (daily 3 times/4 times). Majorly suitable for a) Input for DWH projects b) Can act like enterprise database incase transactional processing database goes down.

OLAP (ONLINE ANALYTICAL PROCESSING)

OLAP is an abbreviation of ONLINE ANALYTICAL PROCESSING. This system is an application that collects, manages, processes and presents multidimensional data for Analysis and management purposes.

DATA MINING

DATA MINING is the process of Analyzing data from different Perspectives and summarizing it into useful information.

BI (BUSINESS INTELLIGENCE)

BI is the Leveraging of Data warehouse to help and make business decisions and Recommendations. Information and data rules engines are leveraged here to help make these decisions along with Statistical Analysis tools and data Mining Tools.

OLCP (ONLINE COMPLEX PROCESSING)

OLCP is typified by a moderate number of rows processed against multiple databases with a response in minutes, not seconds. Filling out an application for instant credit in a store is an OLCP transaction. Another example is getting an instant approval on an interest rate for a loan.

DSS (DECISION SUPPORT SYSTEMS)

Decision Support Systems include batch Reports, which Roll-up numbers to give business the big picture, and overtime, have evolved.

Instead of pre-written scripts, users now require the ability to do adhoc queries, which are unpredictable in their processing and which allow "what if types of questions. These types of questions are essential for long range Planning.

A transaction is a logical unit of work.	TRADITIONAL	Type	Examples	Number of Rows Accessed	Response Time
		OLTP	Update a checking account to reflect a deposit. Debit transaction takes place against current balance to reflect amount of money withdrawn at ATM.	Small	Seconds
	TODAY	DSS	How many child size blue jeans were sold across all of our Eastern stores in the month of March? What were the monthly sales of shoes for retailer X?	Large	Seconds or minutes
		OLCP	Instant credit—How much credit can be extended to this person? What interest rate for a loan can be given to this customer?	Small to moderate against multiple databases	Minutes
		OLAP	Show the top ten selling items across all stores for 1997. Show a comparison of sales from this week to last week.	Large of detail rows or moderate of summary rows	Seconds or minutes

What is a Modern Data Warehouse?

Modern Data warehouses comprised of multiple programs impervious to User. Polyglot persistence encourages the most suitable data storage technology based on data. This "best-fit engineering" aligns multi-structure data into data lakes and considers NoSQL solutions for JSON formats. Pursuing a polyglot persistence dat strategy benefits from virtualization and takes advantage of the different infrastructure. Modern DW requires Petabytes of storage and

more optimized techniques to run complex analytic queries. The traditional methods are relatively less efficient and not cost-effective to fit into the modern day Data Warehousing needs. There are tons of Cloud solutions to build data warehouses performance optimized, inexpensive, and support parallel query execution.

- Incorporate Hadoop, traditional data warehouse, and other data stores.
- Includes multiple repositories may reside in different locations.
- Include Data from mobile devices, sensors, cloud and the Internet of Things.
- Includes structure/semi-structured/unstructured, raw data.
- Inexpensive commodity hardware in cluster mode.

How Modern Data Warehouse Works?

Multiple Parallel Processing (MPP) Architectures

- MPP architecture enables a mighty scale and Distributed Computing.
- Resources add for a linear scale-out to the largest Data Warehousing projects.
- Multiple parallel processing architecture uses a "shared-nothing". There are numerous physical nodes, each runs its instance. This results from performance many times faster than traditional architectures.

Multi-Structured Data

- Define Big Data & Analytics Infrastructure for multiple storage data with a polyglot persistence strategy.
- Integrate portions of the data into the Data Warehouse.
- Federated query access.

Lambda Architecture

In lambda, architecture defines three layers -

- **Speed Layer** - Low latency data.
- **Batch Layer** - Raw Data processing to support complex analysis.
- **Serving Layer** - Response to queries.

Hybrid Architecture

Scale up MPP compute nodes during -

- Peak ETL data loads.
- High query volumes.
- Utilize existing On-Premises data structures.
- Use Cloud services for Advanced Analytics.

Why Modern Data Warehouse Matters?

How Modern Data Warehousing Solves Problems for Businesses -

Data Lakes - Instead of storing in hierarchical files and folders, as traditional data warehouse do, a data lake is the repository that holds a vast amount of raw data in its native format until needed.

Data divided across organizations - Modern Data Warehousing allows for quicker information Assortment and Analysis across organizations and divisions. It keeps the Agility model and promotes more alignment and sooner effect.

IoT streaming data - The Internet of Things has completely transformed the scenario, units, etc. share and stock data across multiple devices.

Business Challenges

- Reduce the cost to store and manage data growth.
- Business demand to analyze new data sources requires investment in technologies to process all data formats.
- Current Data Warehouses good for Multidimensional Analytics but not suited for Image, Video or other new types of analytics.

How to Adopt Modern Data Warehouse?

Growing an Existing DW Environment

- Internal to the Data Warehouse
- Data modeling strategies
- Partitioning
- Clustered columnstore index
- In-memory structure
- MPP

Augment the Data Warehouse

- Complementary Data Storage & Analytical solutions.
- Cloud & Hybrid solutions.
- Data Virtualization/ Virtual DW.

Features of Modern Data Warehouse

- Variety of subject areas & data sources for analysis with the capability to handle the large volume of data.
- Expansion beyond a single relational DW/Data Mart structure to include Data Lake.
- Logical design across multi-platform architecture balancing performance & scalability.

- Data virtualization in addition to Data Integration.
- Support for all type & levels of users.
- Flexible deployment decoupled from the tool used for development.
- Governance model to support security and trust, and Master Data Management.
- Support for promoting the self-service solution to the corporate environment.
- Ability to facilitate Real-Time analysis of high-velocity data.
- Support for Advanced Analytics.
- Agile Delivery approach with the fast delivery cycle.
- Hybrid Integration with Cloud services.
- APIs for downstream access to data.
- Some DW automation to improve speed, consistency, business terminology.
- An analytics sandbox or workbench area to facilitate agility within a BI environment.
- Support for self-service BI to augment corporate BI; Data discovery, Data Exploration, Self-service Data preparation.

Best Practices of Data Warehouse

Define the Compression Formats and Data Storage - There can be more than one option for data storage. Each storage option offers distinct advantages and benefits. It is necessary to evaluate the data formats and storage to work smoothly with the applications in an ecosystem.

Look out for Multi-tenancy Support - Multi-tenancy support is important for the BI environment. It gives the advantage of using a single software stack to serve thousand of partners & customers and make upgrades or customization.

Review the Schema - Evaluate the nature of the database storage. Verify how it's loaded, processes, and analyzed to optimize schema objects.

Ensure Metadata Management - Ensure end-to-end Metadata Management for Data Warehouse initiatives Metadata Management defines. Metadata Management establishes the success of Modern Data Warehousing projects. It captures the necessary information to build, use and interpret the Data Warehouse elements.

Benefits of Modern Data Warehouse

- Rapid integration of data into the environment.
- Improved efficiency in integration reducing time, cost and efforts.
- Opportunity to enable innovative new data models.
- Potential for new insights into the data that provide Preventive analysis and Predictive Analysis.
- Ability to have more extensive datasets for analysis as the data collected and stored continues to grow exponentially.
- Cost advantages of Open source software & Commodity hardware.

Concluding Modern Data Warehouse

The opportunities of Big Data and Advanced analytics are a big challenge. The most sophisticated traditional Data Warehouse changing to meet the requirements of the Modern Data Enterprise. Increase in volume expected to continue. Business velocity continues to change business operations and customer interactions. Data becomes even more diverse and more available than ever before. Big Data means a big impact on business. To dig into the immense new opportunities of Big Data, the Modern enterprise needs a Modern data platform.

[Microsoft Modern Data Warehouse](#) delivers platform, solutions, features, functionality, and benefits that empower the Modern Enterprise in three essential areas -

- Easily manage relational and non-relational data at all volumes and high performance.
- Enjoy a consistent experience across On-premises and Cloud.
- Gain insights from BI and Advanced analytics across all data wherever it resides.

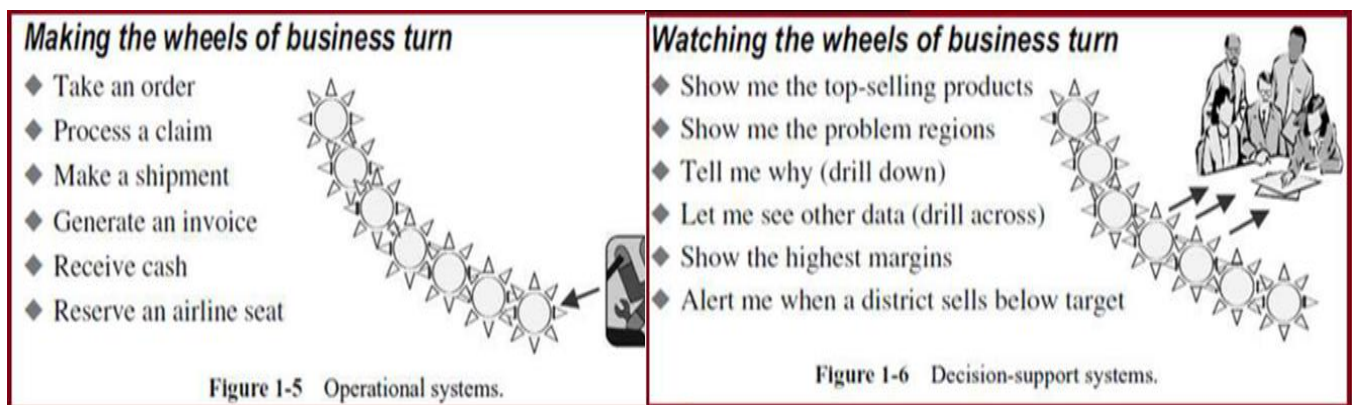
LEGACY DATAWAREHOUSE OTHER DEFINITIONS

Simple Definition:

Database which stores huge volumes of data and mainly designed for decision making.

THE COMPELLING NEED FOR DATA WAREHOUSING

- _ Companies are desperate for strategic information to counter fiercer competition, extend market share, and improve profitability.
 - _ In spite of tons of data accumulated by enterprises over the past decades, every enterprise is caught in the middle of an information crisis. Information needed for strategic decision making is not readily available.
 - _ All the past attempts by IT to provide strategic information have been failures. This was mainly because IT has been trying to provide strategic information from operational systems.
 - _ Informational systems are different from the traditional operational systems. Operational systems are not designed for strategic information.
 - _ We need a new type of computing environment to provide strategic information.
- The data warehouse promises to be this new computing environment.



DIFFERENCES BETWEEN OLTP AND OLAP[DWH]APPLICATIONS

OLTP	DWH
Designed for transaction processing	Designed for Decision support
Volatile	Store Non-volatile
Store current data only	Store historical data
Store detail data	Store summarized data
More joins	Less joins
Normalized	<u>Denormalized</u>
Less indexes	More indexes
Low level managers access	High level managers access

OTHER AUTHORS DIFINITIONS

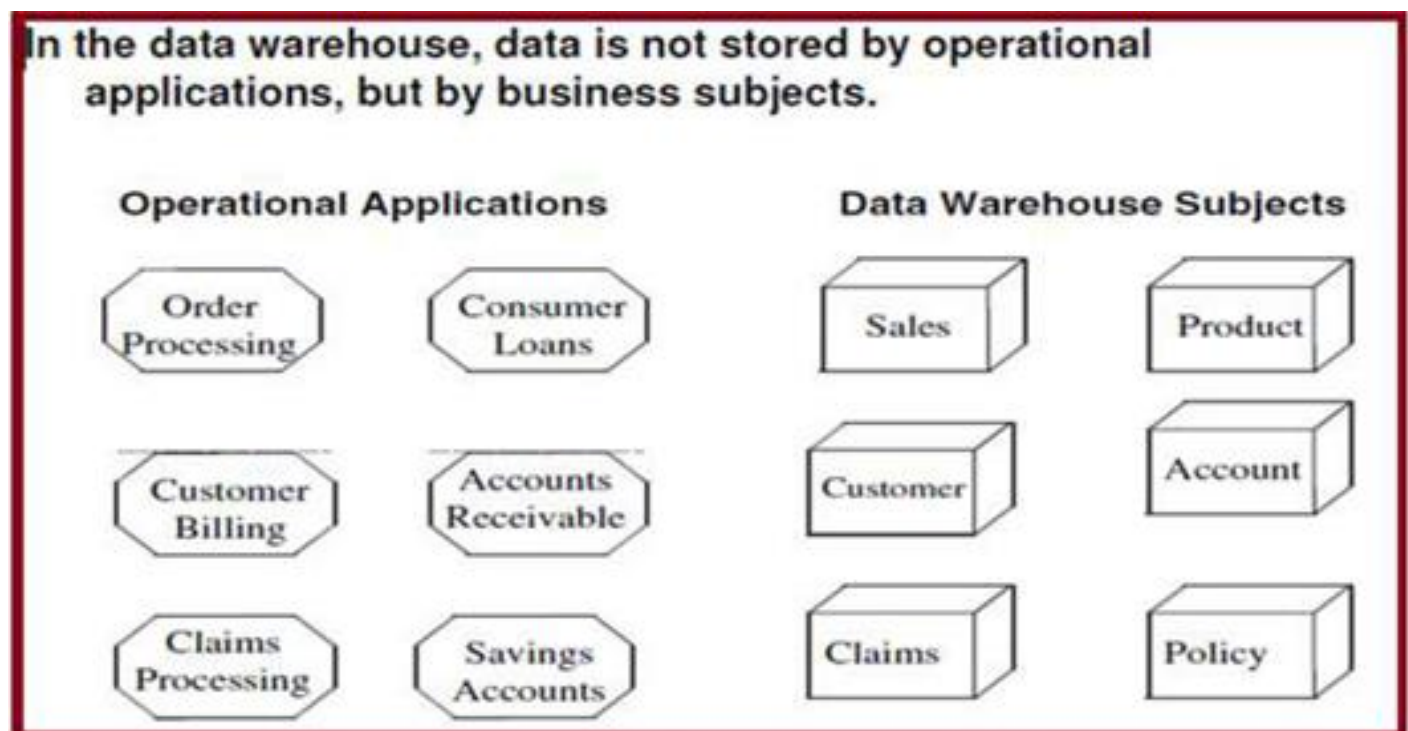
Bill Inmon, considered to be the father of Data Warehousing provides the following

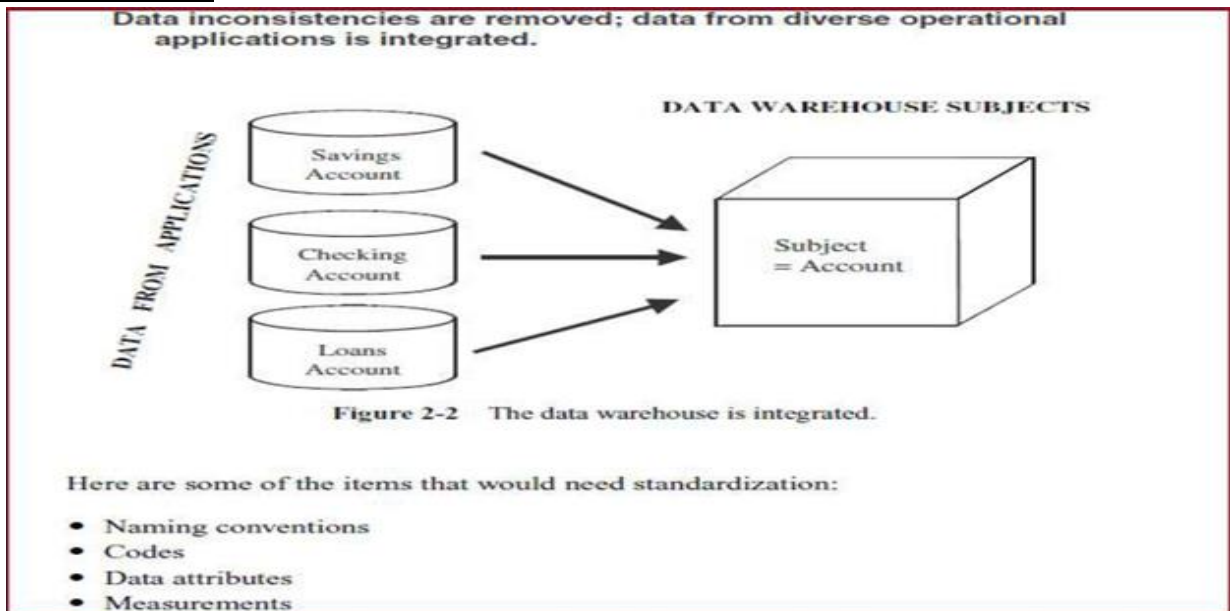
"A Data Warehouse is a subject oriented, integrated, time variant and nonvolatile data"

Sean Kelly, another leading data warehousing practitioner defines the data warehouse in the following way.

The data in the data warehouse is: Separate, Available and Accessible

A)SUBJECT ORIENTED:



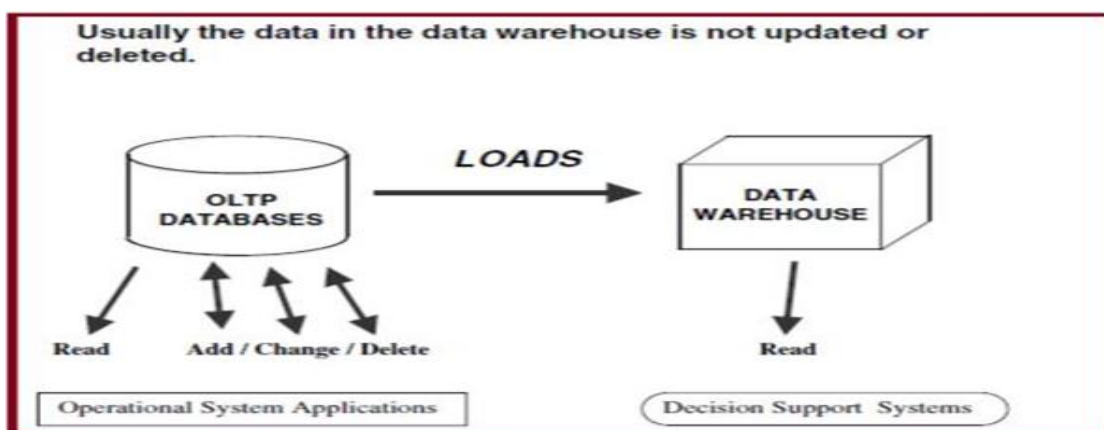
B) INTEGRATED:**C) Time Variant**

The time-variant nature of the data in a data warehouse

- Allows for analysis of the past
- Relates information to the present
- Enables forecasts for the future

D) NONVOLATILE:

Every business transaction does not update the data in the data warehouse. The business transactions update the operational system databases in real time. We add, change, or delete data from an operational system as each transaction happens but do not usually update the data in the data warehouse. You do not delete the data in the data warehouse in real time. Once the data is captured in the data warehouse, you do not run individual transactions to change the data there. Data updates are commonplace in an operational database; not so in a data warehouse. The data in a data warehouse is not as volatile as the data in an operational database is. The data in a data warehouse is primarily for query and analysis.



DATA GRANULARITY

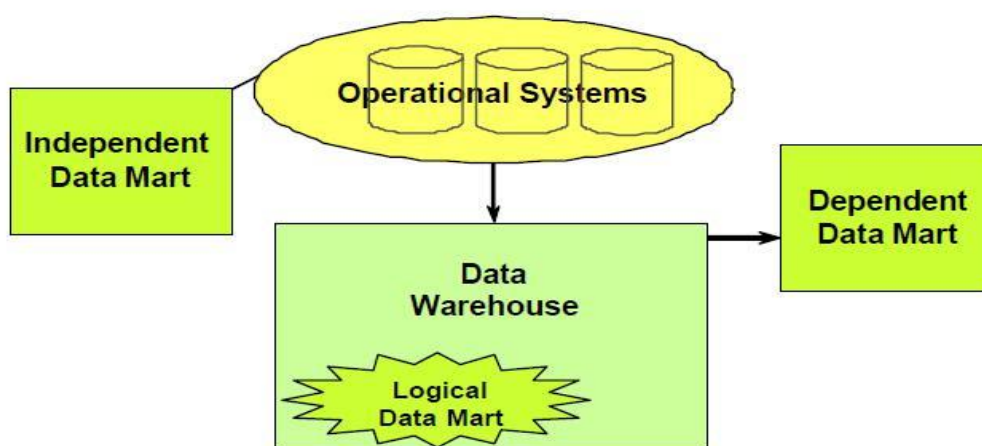
Data granularity in a data warehouse refers to the level of detail data. The lower the level of detail, the finer the data granularity. When you maintain low level of data (day), then upper levels you can analyze easily (week, fortnight, month etc...)

<u>THREE DATA LEVELS IN A BANKING DATA WAREHOUSE</u>		
<u>Daily Detail</u>	<u>Monthly Summary</u>	<u>Quarterly Summary</u>
Account	Account	Account
Activity Date	Month	Month
Amount	Number of transactions	Number of transactions
Deposit/Withdrawal	Withdrawals	Withdrawals
	Deposits	Deposits
	Beginning Balance	Beginning Balance
	Ending Balance	Ending Balance
Data granularity refers to the level of detail. Depending on the requirements, multiple levels of detail may be present. Many data warehouses have at least dual levels of granularity.		

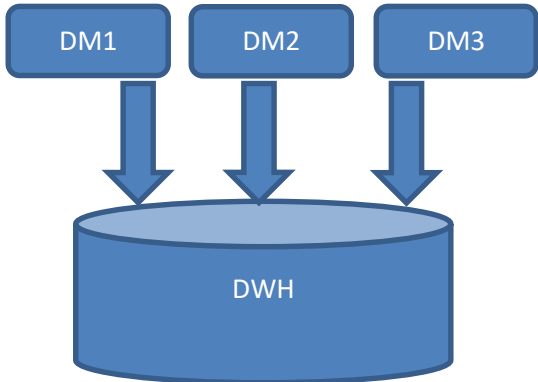
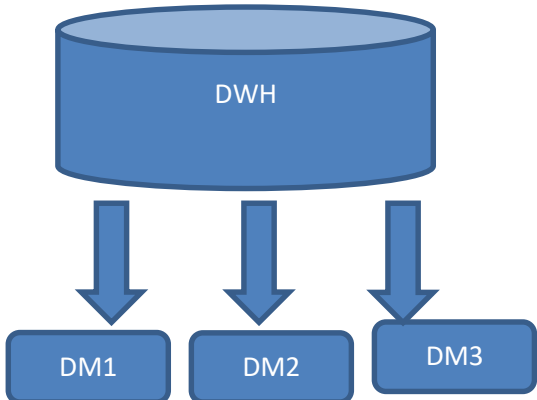
DATA MART AND DATA MART TYPES

A data structure that is optimized for access. It is designed to facilitate end-user analysis of data. It typically supports a single, analytic application used by a distinct set of workers.

- A data mart is a special purpose subset of enterprise data for a particular function or application. It may contain detail or summary data or both.
- Data mart types:
 - Independent—created directly from operational systems to a separate physical data store.
 - Logical—exists as a subset of existing data warehouse.
 - Dependent—created from data warehouse to a separate physical data store.



Practical Way:**DATA WAREHOUSE APPROACHES**

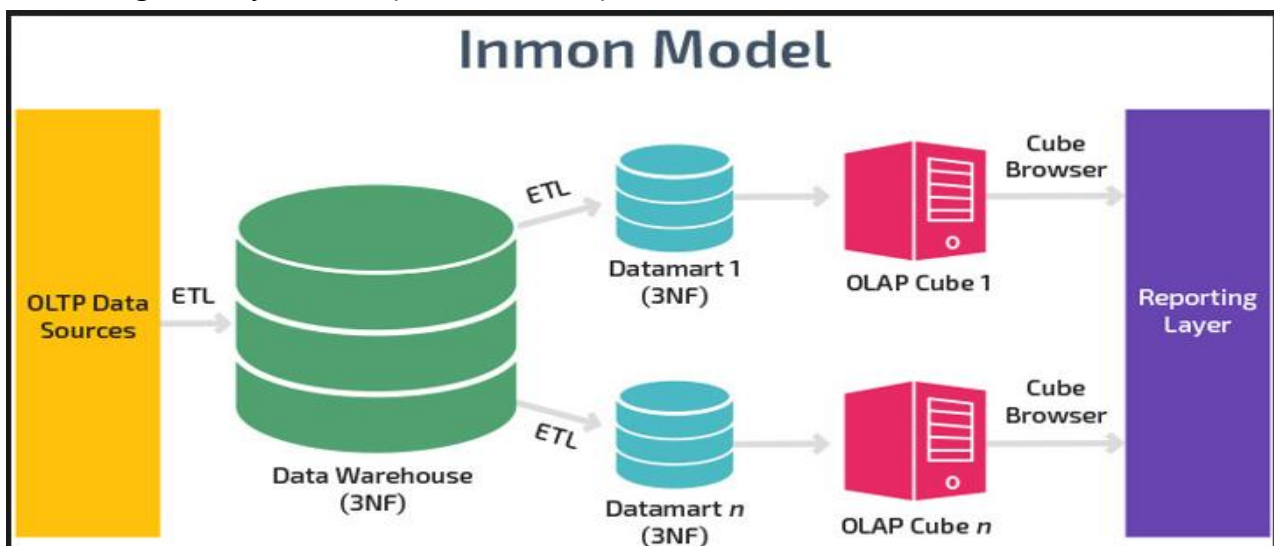
Ralph Kimball (Bottom up approach)	Inmon approach (Top down approach)
<p>Data marts created first and then DWH</p> 	<p>DWH created first and then Data marts</p> 

Top-Down Approach[INMON]**The advantages of this approach are:**

- A truly corporate effort, an enterprise view of data
- Inherently architected not a union of disparate data marts
- Single, central storage of data about the content centralized rules and control
- May see quick results if implemented with iterations

The disadvantages are:

- Takes longer to build even with an iterative method
- High exposure/risk to failure
- Needs high level of cross-functional skills
- High outlay without proof of concept

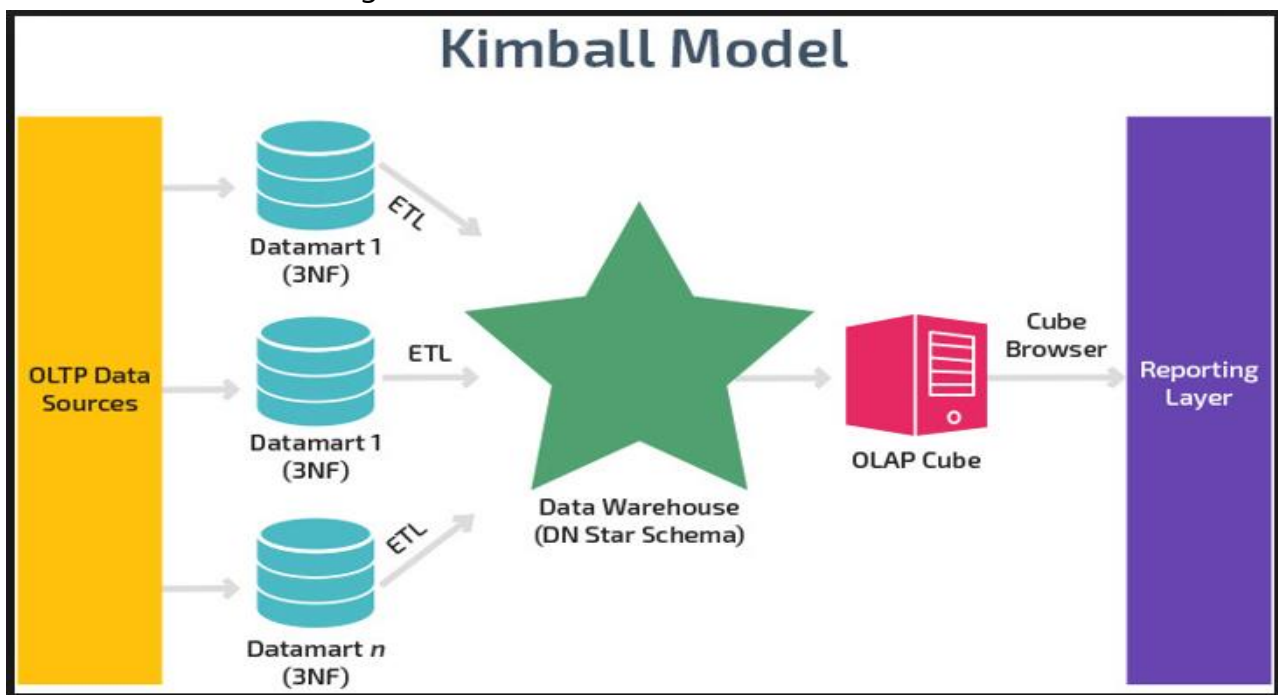


Bottom-Up Approach[KIMBALL]**The advantages of this approach are:**

- Faster and easier implementation of manageable pieces
- Favorable return on investment and proof of concept _ Less risk of failure
- Inherently incremental; can schedule important data marts first
- Allows project team to learn and grow

The disadvantages are:

- Each data mart has its own narrow view of data _ Permeates redundant data in every data mart
- Perpetuates inconsistent and irreconcilable data
- Proliferates unmanageable interfaces



Qs: Which approach your company is using in your project?

DWH LIFE CYCLE AND STEPS:

It is the combination of 3 life cycles [ETL, Semantic and Reporting]

Common Steps in all Life Cycles.

- (1) Requirements Gathering
- (2) Analyzing the requirements
- (3) Designing the business
- (4) Coding the business
- (5) Testing the business
- (6) Implementation / Deployment
- (7) Maintenance (warranty support / SLA support (Service Level Agreement))

OLAP TYPES (MLOAD, ROLAP, DOLAP, HOLAP)**(1)ROLAP: - (Relational OLAP)**

Applied on relational sources, both data and aggregate information stored in relational sources

Ex: BO, Congo's, Crystal reports, Micro Strategy

(2)MOLAP: - (Multidimensional OLAP)

Here, Analysis will be done in multidimensional applications.

Here, data and aggregate information stored under multidimensional sources

(3)HOLAP: - (Hybrid OLAP)

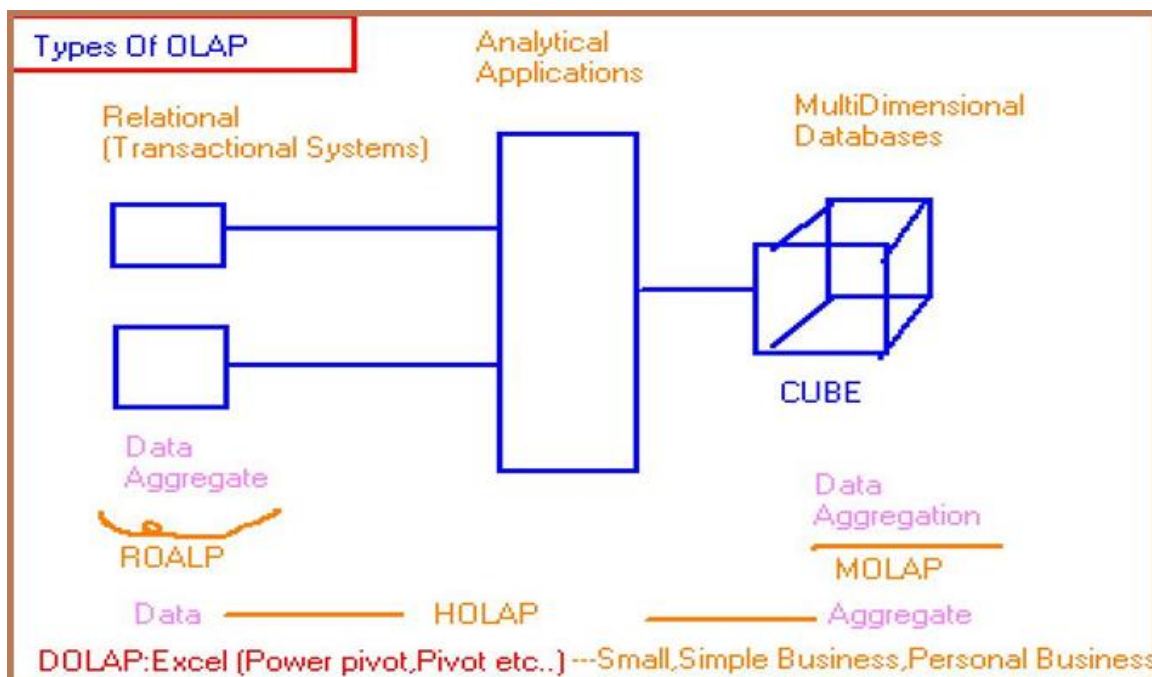
Here data stored in relational sources and aggregated values under multidimensional sources such as " cubes"

Ex: Cognos, BO, Micro Strategic ...

(4)DOLAP:- (Desktop OLAP)

Here, the analysis will be done in desktop applications

Ex: MS-access, Excel, Lotus etc.

Picture

DATA SCRUBBING, CLEANSING, PROFILING, MERGING & TRANSFORMING**DWH FAQS****Q: Differences between OLTP and OLAP?**

IT applications

OLTP [Online Transaction Processing]

Purpose: Day-Day business storage

Process: Front-End, Back-End process

Points:

- a) Data: Less data
- b) Model: Normalized model (more tables, smaller tables)
- c) Joins: More joins (So retrieval it is slower)
- d) Access Frequency: High [Milliseconds to seconds]
- e) Volatility: Volatile [Read, Insert, update, Delete]

2) OLAP [Online analytical Processing]

Purpose: For analysis and mining

Process: Layered architecture

Points:

- a) Data: More data
- b) Model: Denormalized model (Less tables, complex tables)
- c) Joins: Less joins (So retrieval it is faster)
- d) Access Frequency: Low [Day/week/month...]
- e) Volatility: Nonvolatile [Read]

Q: What is Data warehouse [Single line] and warehouses available in the market?

A: Huge storage area which is suitable for decision making.

Ex: Teradata, Oracle, Sql Server, Sybase, Netezza, GreenPlum etc...

Q: Difference between OLAP and BI?

A: BI is a process which uses OLAP approach to gather, convert and presents the data.

Q: Difference between Analysis and Mining?

A: Analysis talks about 'current understanding of data' and 'past analysis of data', whereas mining talks about 'future prediction of data'.

Note: Data mining is a knowledge analysis and discovery method.

Q: How many types of decision making available in IT?

A:

- a) Strategic-- Past analysis [OLAP data]
- b) Tactical--Current and recent [OLTP]

Q: Difference between OLTP and ODS?

A:

OLTP holds real-time data whereas ODS holds near real time data [Ex:daily 4 times from OLTP]

ODS helps in two ways

- a) Input to DWH
- b) In case enterprise database [OLTP] fails, it acts like enterprise db and continue the Operations.

Note: All top companies use ODS in the ETL flow.

Q: OLTP similar terminologies?

A:

- a) Operational Systems
- b) Enterprise databases
- c) Business Processing Systems

d) Transactional systems

Q: Can you describe an End-End typical ETL Flow?

A:

OLTP---->ODS---->Stage Area [Files]---->DWH[Stage]---->DWH[WorkArea]---->DWH[Target]

Q: How many life cycles does a DWH project need?

A: Three life cycles a) ETL b) Semantic c) Reporting

Famous DWH life cycles: Agile, iterative incremental [Block approach], V model

Q: Explain DWH characteristics?

Inman Characteristics:

a) **Subject oriented**: Store data based on operation but not application.

Ex: Savings count subject area

Application2: a) Withdraw b) Deposit

Subject Area: withdraw+deposit single area

b) **Integrated**: Integrate from diverse applications by eliminating inconsistencies and following standards.

c) **Nonvolatile**: No change of data. [Don't modify the data]

Helps to maintain history.

d) **Time variant**: Data store base on timeframes [granularity load / grain]

We can perform the below operations easily.

a) Current understanding

b) Past analysis

c) Future predictions

Q: What is granularity and which grain your project has?

A:

Granularity talks about the level of detailed data maintaining in the project.

Dwh has multiple tables with multiple grains.

Always the lowest grain has lot of burden but improves analytical performance.

Ex: Day/ week /month etc... Grains

Day grain is recommended for better analysis.

Q: What is datamart and how many types available?

A:

DataMart--Data sub store for specific business/ operation / functionality

May or not be a subset of DWH.

Three types of datamarts

- a) **Dependent**-- DWH created first and then datamart [**subset of DWH**]

Ex: ICICI DWH, Savings account dependent datamart

- b) **Independent**-- Directly created from source systems [we don't use DWH here]

Ex: ICICI employee payroll System [No history, not much analysis required]

- c) **Logical Datamart**--It is replica of another data mart [Holds structure of datamart]

Q: How do we create datamarts practically?

A:

- a) By using Complex Views [**Materialized views**] --Dependent DM

- b) By using a complex table --Dependent DM

- c) By using a separate physical storage area --Independent DM

Q: How many types of DWH approaches available, which approach your company following?

As:

Two approaches

- a) **Kimball** approach--Datamarts-> DWH [**Bottom up approach**]

- b) **Inman** approach-- DWH -->Datamarts [**Top down approach**]

My company using Kimball approach for DWH implementation because of dynamic decisions and adhoc requests.

Q: How many types of OLAP available, which is effective?

A:

OLAP--Online Analytical Processing

It create aggregates (sum, avg, max, min, stddev, covariance etc...) for decision making.

ROLAP: Relational OLAP--Data and aggregates in the relational area [OLTP area]

Latency time less [fresh data analysis]

Slower analysis

Minimal set up

Ex: Small, Medium

MLOAP: Multidimensional OLAP-- Data and aggregates in multidimensional area [Cube]

Adv: More and detailed analysis this is helpful

Latency time is high [fresh data analysis is not possible]

Complex set up

Ex: Corporates

HOLAP: Data in relational area and aggregates in multidimensional area [cube]

Ex: Latency time average

Avg set up

DOLAP: Desktop OLAP: Here the data and aggregates on PC based desk top applications

Ex: Excel, Lotus, FoxPro, VISICALC etc...

Q: Explain Date hierarchy?

A: Day-->Week-->Fort Night-->Month-->Quarter-->Half Year (Semester)-->Year→Quad year→Decade→Century

Q: Explain Time Hierarchy?

A: milliseconds→Seconds→Minutes→Hours

Q: You have two databases with same size, then how do you identify which is OLTP or OLAP?

A: Normalized model is OLTP, DE normalized model is OLAP.

Q: Choose the below are required to implement a DWH Project?

A:

- a) Life Cycle
- b) Approach
- c) Data Model
- d) Type of OLAP
- e) Characteristics / principles
- f) Historical data

Correct; a, b, c, e

Q: What is Self Service BI?

A:

If an application support Personal, Team and Corporate analysis, then that application is SSBI application.

With respect to Power BI, SSBI indicate accessing, creation, managing {content packs and apps}, reports and dash boards.

Note: Users who have Pro account, they consider as SSBI professionals.

Q: What is OLTP?

A: Online Transaction Processing, this process helps to store daily business data.

Q: What is OLAP?

A: Online Analysis Processing, this process gathers, converts and presents data.

Q: What is BI?

A:

It is the process of gathering, converting and presenting business data for decision making purpose.

Q: What is Data warehouse? [Business data]

A:

Data Huge Storage Area.

Q: What is a Modern Data Warehouse? [Social media, IOT, Business data and Cloud data {Lake}]

Modern Data warehouses comprised of multiple programs impervious to User.

Polyglot persistence encourages the most suitable data storage technology based on data. This "best-fit engineering" aligns multi-structure data into data lakes and considers NoSQL solutions for JSON formats. Pursuing a polyglot persistence data strategy benefits from virtualization and takes advantage of the different infrastructure. Modern DW requires Petabytes of storage and more optimized techniques to run complex analytic queries.

Q: How many types of Data warehouses available?

A:

- a) Structured storage Data warehouse [Databases used for this purpose]
- b) Unstructured storage Data warehouse [File Streams and Cloud Storages used]

Ex: Data Lake

Q: What is Internet of things [IoT]?

A:

The Internet of things (IoT) refers to the concept of extending Internet connectivity beyond conventional computing platforms such as personal computers and mobile devices, and into any range of traditionally "dumb" or non-internet-enabled physical devices and everyday objects. Embedded with electronics, Internet connectivity, and other forms of hardware (such as sensors), these devices can communicate and interact with others over the Internet, and they can be remotely monitored and controlled

Q: What is Data Insight?

A: The capacity to gain an accurate and deep understanding of someone or something

Q: What do you mean by Analytics?

A: The systematic computational analysis of data or statistics.

Information resulting from the systematic analysis of data or statistics.

"these analytics can help you decide if it's time to deliver content in different ways"

Analytics is the scientific process of discovering and communicating the meaningful patterns which can be found in data.

Q: What is Polybase?

A:

PolyBase is a technology that accesses and combines both non-relational and relational data, all from within SQL Server. It allows you to run queries on external data in Hadoop or Azure blob storage. The queries are optimized to push computation to Hadoop.

Q: What is Hadoop?

Hadoop is an open source distributed processing framework that manages data processing and storage for big data applications running in clustered systems. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs.

Q: What is Azure?

A:

Microsoft Azure is a cloud computing service created by Microsoft for building, testing, deploying, and managing applications and services through Microsoft-managed data centers.

Q: What can Microsoft Azure Do?

Microsoft maintains a growing directory of Azure services, with more being added all the time. All the elements necessary to build a virtual network and deliver services or applications to a global audience are available, including:

Virtual machines. Create Microsoft or Linux virtual machines (VMs) in just minutes from a wide selection of marketplace templates or from your own custom machine images. These cloud-based VMs will host your apps and services as if they resided in your own data center.

SQL databases. Azure offers managed SQL relational databases, from one to an unlimited number, as a service. This saves you overhead and expenses on hardware, software, and the need for in-house expertise.